Biometrika (2013), **100**, 4, *pp.* 781–800
Printed in Great Britain
doi: 10.1093/biomet/ast020
Advance Access publication 29 July 2013

Bridging the ensemble Kalman and particle filters

By M. FREI and H. R. KÜNSCH
Seminar für Statistik, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
frei@stat.math.ethz.ch kuensch@stat.math.ethz.ch

Summary

In many applications of Monte Carlo nonlinear filtering, the propagation step is computationally expensive, and hence the sample size is limited. With small sample sizes, the update step becomes crucial. Particle filtering suffers from the well-known problem of sample degeneracy. Ensemble Kalman filtering avoids this, at the expense of treating non-Gaussian features of the forecast distribution incorrectly. Here we introduce a procedure that makes a continuous transition indexed by $\gamma \in [0, 1]$ between the ensemble and the particle filter update. We propose automatic choices of the parameter γ such that the update stays as close as possible to the particle filter update subject to avoiding degeneracy. In various examples, we show that this procedure leads to updates that are able to handle non-Gaussian features of the forecast sample even in high-dimensional situations.

Some key words: Ensemble Kalman filter; Nonlinear filtering; Particle filter; State space model.

1. Introduction

State space models consist of a discrete or continuous time Markov process that is partially observed at discrete time-points and subject to independent random errors. Estimation of the state at time t given observations up to the same time is called filtering or data assimilation. Since exact computations are possible essentially only in linear Gaussian situations, Monte Carlo methods are typically used for filtering. In many environmental applications, in particular in atmospheric physics, oceanography and reservoir modelling, the dimension of the state is very large, and the computational costs of simulation from the state transitions are huge, which severely limits the potential sample size for Monte Carlo filtering methods. Standard particle filters (Gordon et al., 1993; Pitt & Shephard, 1999; Doucet et al., 2000) suffer from sample degeneracy (Snyder et al., 2008). In contrast, the ensemble Kalman filter (Evensen, 1994; Burgers et al., 1998; Houtekamer & Mitchell, 1998) can handle some problems where the dimensions of states and observations are large, and the number of replicates is small, but at the expense of incorrectly treating non-Gaussian features of the forecast distribution that arise in nonlinear systems.

To relax the Gaussian assumption, two paradigms are predominant: mixture filters that approximate the forecast distribution as a mixture of Gaussian distributions (Bengtsson et al., 2003; Sun et al., 2009; Dovera & Della Rossa, 2011; Stordal et al., 2011; Hoteit et al., 2012; Rezaie & Eidsvik, 2012; Frei & Künsch, 2013), and sequential importance samplers that use the ensemble Kalman filter as a proposal distribution (Mandel & Beezley, 2009; Papadakis et al., 2010). In this article, we introduce an update scheme that blends these: a Gaussian mixture proposal obtained from an ensemble Kalman filter update based on a tempered likelihood is corrected by a particle filter update. In this way we do not have to fit a Gaussian mixture to the

forecast sample or to approximate the ratio of the predictive density to the proposal density. A further advantage of our procedure is that we can implement these two steps in such a way that the particle weights do not depend on artificial observation noise variables and the resampling avoids ties. Our method does not require any special structure for the state dynamics: they may be stochastic or deterministic, all we need is the ability to simulate from the state transitions. This contrasts with some of the recent generalizations to particle filtering as proposed in, e.g., van Leeuwen (2010) or Morzfeld & Chorin (2012).

Our procedure depends on a single tuning parameter $\gamma \in [0, 1]$, which allows continuous interpolation between the ensemble Kalman filter and the particle filter. Hence, the parameter γ controls the bias-variance trade-off between a correct update and maintaining the diversity of the sample. It can be chosen without prior knowledge based on a suitable measure of diversity like effective sample size (Liu, 1996), or the expected number of Gaussian components that are represented in the resample.

2. Problem setting, notation and background material

We consider a dynamical system with state variables $\{x_t \in \mathbb{R}^q : t = 0, 1, \ldots\}$ and observations $\{y_t \in \mathbb{R}^r : t = 1, 2, \ldots\}$. The state follows a deterministic or stochastic Markovian evolution, that is $x_t = g(x_{t-1}, \xi_t)$, where the system noise ξ_t is independent of all past values x_s and all $\xi_s, s < t$. There is no need to know the function g in explicit form, we assume only that for given x_{t-1} we are able to simulate from the distribution of $g(x_{t-1}, \xi_t)$. In particular, the evolution can be in continuous time, given by an ordinary or stochastic differential equation.

In all cases we assume linear observations with Gaussian noise: $y_t = Hx_t + \epsilon_t$, where $\epsilon_t \sim \mathcal{N}(0, R)$. Thus the likelihood for the state x_t given the observation y_t is $\ell(x_t \mid y_t) = \varphi(y_t; Hx_t, R)$. Here and below $\varphi(x; \mu, \Sigma)$ denotes the multivariate normal density at x with mean μ and covariance Σ. In the final section, we will discuss briefly how to modify the method for non-Gaussian likelihoods.

We denote all observations up to time t, $(y_1, \ldots, y_t)$, by $y_{1:t}$. The forecast distribution π_t^p at time t is the conditional distribution of x_t given $y_{1:t-1}$, and the filter distribution π_t^u at time t is the conditional distribution of x_t given $y_{1:t}$. In principle, these distributions can be computed recursively, alternating between propagation and update steps. The propagation step leads from π_{t-1}^u to π_t^p: π_t^p is the distribution of $g(x_{t-1}, \xi_t)$, where $x_{t-1} \sim \pi_{t-1}^u$ and ξ_t is independent of x_{t-1} and has the distribution given by the evolution of the system. The update step leads from π_t^p to π_t^u and is Bayes's formula: $\pi_t^u(\mathrm{d}x_t) \propto \ell(x_t \mid y_t)\pi_t^p(\mathrm{d}x_t)$. However, analytical computations are possible essentially only if the system evolution is also linear with additive Gaussian noise. Hence one resorts to Monte Carlo approximations, i.e., one represents π_t^p and π_t^u by ensembles $\{x_{t,j}^p\}$ and $\{x_{t,j}^u\}$ respectively. The members of these ensembles are called particles.

The propagation step lets the particles evolve according to the dynamics of the state, i.e., we simulate according to the time evolution starting at $x_{t-1,j}^u$ at time $t-1$, $x_{t,j}^p = g(x_{t-1,j}^u, \xi_{t,j})$. However, the computational complexity of this step limits the number N of particles, that is the size of the sample.

The bootstrap particle filter (Gordon et al., 1993) updates the forecast particles by weighting with weights proportional to the likelihood $\ell(x_t \mid y_t)$ and converts this into an unweighted sample by resampling, i.e., $\{x_{t,j}^u\}$ is obtained by sampling from

$$\sum_{i=1}^{N} \omega_{t,i} \Delta_{x_{t,i}^p}, \quad \omega_{t,i} = \frac{\ell(x_{t,i}^p \mid y_t)}{\sum_{j=1}^{N} \ell(x_{t,j}^p \mid y_t)}.$$

Thus some of the forecast particles disappear and others are replicated. If the likelihood is quite peaked, which is the case in high dimensions with many independent observations, the weights will be heavily unbalanced, and the filter sample eventually degenerates since it concentrates on a single or a few particles; see Snyder et al. (2008). Auxiliary particle filters (Pitt & Shephard, 1999) can attenuate this behaviour to some extent, but they require good proposal distributions for the propagation, and an analytical expression for the transition densities.

The ensemble Kalman filter (Burgers et al., 1998; Houtekamer & Mitchell, 1998) makes an affine correction of the forecast particles based on the new observation y_t and artificial observation noise variables $\epsilon_{t,j} \sim \mathcal{N}(0, R)$:

$$x^u_{t,j} = x^p_{t,j} + K(\hat{P}^p_t)\left(y_t - Hx^p_{t,j} + \epsilon_{t,j}\right),$$

where $\hat{P}^p_t$ is an estimate of the forecast covariance at time t, typically a regularized version of the sample covariance of $\{x^p_{t,j}\}$, and $K(P)$ is the Kalman gain $K(P) = PH'(HPH' + R)^{-1}$. This update formula is asymptotically correct under the assumption that the forecast distribution π^p_t is Gaussian (Le Gland et al., 2011). Although this is usually not true, the update nevertheless has been found to work well in a variety of situations; see Evensen (2007) and references therein. For later use, we note that conditional on the forecast sample,

$$x^u_{t,j} \sim \mathcal{N}\{x^p_{t,j} + K(\hat{P}^p_t)(y_t - Hx^p_{t,j}),\, K(\hat{P}^p_t)RK(\hat{P}^p_t)'\}.$$

Therefore the filter sample can be considered as a balanced sample from the conditional Gaussian mixture

$$\frac{1}{N}\sum_{i=1}^{N} \mathcal{N}\{x^p_{t,i} + K(\hat{P}^p_t)(y_t - Hx^p_{t,i}),\, K(\hat{P}^p_t)RK(\hat{P}^p_t)'\}. \tag{1}$$

Here, balanced sample simply means that we draw exactly one realization from each of the N equally weighted Gaussian components. One can also adjust the first two moments systematically by a deterministic affine correction of the forecast particles, which leads to the class of so-called ensemble square root filters (Anderson, 2001; Whitaker & Hamill, 2002; Tippett et al., 2003). Here, we focus on the stochastic version, for two reasons: first, the representation (1) is crucial for our developments; second, the stochastic ensemble Kalman filter is known to be more robust than the deterministic variants in nonlinear and/or non-Gaussian situations that are of interest to us (Lawson & Hansen, 2004; Lei et al., 2010).

3. Ensemble Kalman particle filter

3·1. *New method*

We consider here the update at a single fixed time t and thus suppress t in the notation. We follow the progressive correction idea (Musso et al., 2001) and write

$$\pi^u(\mathrm{d}x) \propto \ell(x \mid y)^{1-\gamma}\pi^{u,\gamma}(\mathrm{d}x), \quad \pi^{u,\gamma}(\mathrm{d}x) \propto \ell(x \mid y)^{\gamma}\pi^{p}(\mathrm{d}x)$$

where $0 \leqslant \gamma \leqslant 1$ is arbitrary. Our approach is to use an ensemble Kalman filter update to go from π^p to $\pi^{u,\gamma}$, and a particle filter update to go from $\pi^{u,\gamma}$ to π^u. The rationale behind this two-stage procedure is to achieve a compromise between sample diversity and systematic error due to non-Gaussian features of π^p. The former is large if γ is close to one because the ensemble Kalman filter update draws the particles closer to the observation y, and the exponent $1 - \gamma$ dampens the ratio of any two resampling probabilities. The latter is small if γ is small.

Since $\ell(x \mid y)^\gamma \propto \varphi(y; Hx, R/\gamma)$, and

$$PH'(HPH' + R/\gamma)^{-1} = \gamma PH'(\gamma HPH' + R)^{-1} = K(\gamma P), \tag{2}$$

the ensemble filter update is straightforward: we need compute only the gain with the reduced covariance $\gamma \hat{P}^p$. The particle update then resamples with weights proportional to $\ell(x \mid y)^{1-\gamma} \propto \varphi\{y; Hx, R/(1-\gamma)\}$. However, there are two immediate drawbacks to such an algorithm: the particle weights depend on the artificial observation noise variables needed for the ensemble Kalman filter update, and the resampling introduces tied values. We show next how to address both points. By (1), we can write

$$\pi^{u,\gamma} \approx \pi^{u,\gamma}_{\text{EnKF}} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{N}\{v_i^{u,\gamma}, Q(\gamma, \hat{P}^p)\} \tag{3}$$

where

$$v_i^{u,\gamma} = x_i^p + K(\gamma \hat{P}^p)(y - Hx_i^p), \tag{4}$$

$$Q(\gamma, \hat{P}^p) = \frac{1}{\gamma} K(\gamma \hat{P}^p) R K(\gamma \hat{P}^p)'. \tag{5}$$

Instead of sampling from (3) and applying a particle correction, we delay the ensemble Kalman filter sampling step, and update (3) analytically. This is easy because the update of a Gaussian mixture by a Gaussian likelihood is again a Gaussian mixture whose parameters can be computed easily (Alspach & Sorenson, 1972). We obtain

$$\pi^u \approx \pi^u_{\text{EnKPF}} = \sum_{i=1}^{N} \alpha_i^{u,\gamma} \mathcal{N}(\mu_i^{u,\gamma}, P^{u,\gamma}) \tag{6}$$

where EnKPF stands for ensemble Kalman particle filter and

$$\begin{aligned} \alpha_i^{u,\gamma} &\propto \varphi\{y; Hv_i^{u,\gamma}, HQ(\gamma, \hat{P}^p)H' + R/(1-\gamma)\}, \\ \mu_i^{u,\gamma} &= v_i^{u,\gamma} + K\{(1-\gamma)Q(\gamma, \hat{P}^p)\}(y - Hv_i^{u,\gamma}), \\ P^{u,\gamma} &= [I - K\{(1-\gamma)Q(\gamma, \hat{P}^p)\}H]Q(\gamma, \hat{P}^p). \end{aligned} \tag{7}$$

The update consists now in sampling from (6). The mixture proportions $\{\alpha_i^{u,\gamma}\}$ do not depend on the artificial observation noise variables, and even if one $\alpha_i^{u,\gamma}$ dominates, there is still some diversity in the filter sample because the covariance $P^{u,\gamma}$ is not zero if $\gamma > 0$.

Sampling from the ith component of (6) can be done as follows: let ϵ_1 and ϵ_2 be two independent $\mathcal{N}(0, R)$ random variables. Then

$$x^{u,\gamma} = x_i^p + K(\gamma \hat{P}^p)(y + \gamma^{-1/2}\epsilon_1 - Hx_i^p) = v_i^{u,\gamma} + K(\gamma \hat{P}^p)\gamma^{-1/2}\epsilon_1$$

clearly has distribution $\mathcal{N}\{v_i^{u,\gamma}, Q(\gamma, \hat{P}^p)\}$, and thus by standard arguments

$$x^u = x^{u,\gamma} + K\{(1-\gamma)Q(\gamma, \hat{P}^p)\}\{y + (1-\gamma)^{-1/2}\epsilon_2 - Hx^{u,\gamma}\}$$

is a sample from $\mathcal{N}(\mu_i^{u,\gamma}, P^{u,\gamma})$. Hence there is no need to compute a square root of $P^{u,\gamma}$.

A summary of the ensemble Kalman particle filter is given in Algorithm 1.

Algorithm 1. Ensemble Kalman particle filter.
Given a forecast ensemble $\{x_j^p\}$ and an observation y:

Step 1. Compute the estimated forecast covariance $\hat{P}^p$.

Step 2. Choose γ and compute $K(\gamma \hat{P}^p)$ according to (2) and $v_i^{u,\gamma}$ according to (4).

Step 3. Compute $Q(\gamma, \hat{P}^p)$ according to (5) and $\alpha_i^{u,\gamma}$ according to (7).

Step 4. Choose indices $I(j)$ by sampling from the weights $\{\alpha_i^{u,\gamma}\}$ with some balanced sampling scheme; e.g., equation (12) in Künsch (2005).

Step 5. Generate $\epsilon_{1,j} \sim \mathcal{N}(0, R)$ and set $x_j^{u,\gamma} = v_{I(j)}^{u,\gamma} + K(\gamma \hat{P}^p)\gamma^{-1/2}\epsilon_{1,j}$.

Step 6. Compute $K\{(1-\gamma)Q(\gamma, \hat{P}^p)\}$, generate $\epsilon_{2,j} \sim \mathcal{N}(0, R)$ and set $x_j^u = x_j^{u,\gamma} + K\{(1-\gamma)Q(\gamma, \hat{P}^p)\}\{y + (1-\gamma)^{-1/2}\epsilon_{2,j} - Hx_j^{u,\gamma}\}$.

Because matrix inversion is continuous, it is easy to check that as $\gamma \to 0$, $v_i^{u,\gamma} \to x_i^p$, $Q(\gamma, \hat{P}^p) \to 0$, $\alpha_i^{u,\gamma} \to \varphi(y; Hx_i^p, R)$, $\mu_i^{u,\gamma} \to x_i^p$ and $P^{u,\gamma} \to 0$. Hence in the limit $\gamma \to 0$, we obtain the particle filter update. Similarly, in the limit $\gamma \to 1$ we obtain the ensemble Kalman filter update because for $\gamma \to 1$, $\{HQ(\gamma, \hat{P}^p)H' + R/(1-\gamma)\}^{-1}$ converges to zero and thus $\alpha_i^{u,\gamma} \to 1/N$. The ensemble Kalman particle filter therefore provides a continuous interpolation between the particle and the ensemble Kalman filters.

4. Modifications in high dimensions

4·1. *Efficient implementation*

We focus here on the situation where $N \ll q$ with N small and q very large. If the linear algebra is carefully implemented, the ensemble Kalman particle filter has the same asymptotic complexity as the ensemble Kalman filter. The choice of an appropriate implementation depends mainly on the dimension r of the observations. If $r = \mathcal{O}(N)$, the ensemble Kalman particle filter should be implemented as described in Algorithm 1. If $r \gg N$, efficient subspace pseudo-inversion as detailed in Evensen (2007, §§ 14.2 and 14.3) should be adopted. If R is diagonal or block-diagonal, a third possibility is to assimilate the observations one at a time or in small groups using again the implementation as provided in Algorithm 1. For large r, this is more costly than subspace pseudo-inversion but no low-rank approximations are needed.

4·2. *Covariance tapering and localization*

If the dimension q of the state space is not small compared to the number N of particles, the variability of the usual sample covariance is large, and regularization may be necessary. In the context of the ensemble Kalman filter, two regularization techniques are common. Either one uses a tapered estimate for $\hat{P}^p$ (Houtekamer & Mitchell, 2001; Furrer & Bengtsson, 2007), that is, the sample covariance matrix is multiplied elementwise by a sparse correlation matrix that is zero when the distance between two components of the state is larger than some threshold; or one uses localized updates in grid space (Ott et al., 2004; Sakov & Bertino, 2011), where, in the

simplest case, each state variable is updated only with local observations that depend on spatially close state variables in some local window.

As pointed out in, e.g., Lei & Bickel (2011), local update schemes are incompatible with resampling since it is not clear how to glue together the locally updated ensembles to a single global ensemble without introducing discontinuities. For this reason, we focus on covariance tapering, which in principle is straightforward to apply in the context of the ensemble Kalman particle filter. In practice though, tapering has implications for the efficient implementations described above. The direct and serial assimilation schemes are unaffected and can be carried out as usual, but subspace pseudo-inversion is no longer possible since it requires a low-rank $\hat{P}^p$. For large r, a different approach is needed. If the error covariance matrix R and the observation matrix H are sparse, tapering has the additional benefit that the computation of $v_i^{u,\gamma}$ in Step 2 and of $x_j^{u,\gamma}$ in Step 5 is much faster because we do not need to compute $K(\gamma \hat{P}^p)$ for this. It is sufficient to solve $2N$ equations of the form $(\gamma HPH' + R)x = b$, which is fast for sparse matrices. However, this advantage is lost because for $Q(\gamma, \hat{P}^p)$ we need $K(\gamma \hat{P}^p)$, which is in general a full matrix. We could multiply the gain matrix by another taper in order to facilitate the computation of $Q(\gamma, \hat{P}^p)$ and to turn $Q(\gamma, \hat{P}^p)$ into a sparse matrix. This would then make Steps 3 and 6 in Algorithm 1 faster because again all we need to do is to solve $2N$ equations of the form $\{(1-\gamma)HQ(\gamma, \hat{P}^p)H' + R\}x = b$.

Because $K(\gamma \hat{P}^p)$ is used only to compute $Q(\gamma, \hat{P}^p)$, a simpler alternative that avoids computing gain matrices is to generate the values $K(\gamma \hat{P}^p)\gamma^{-1/2}\epsilon_{1,j}$ needed in Step 5 before Step 3 and 4, and then to replace $Q(\gamma, \hat{P}^p)$ by a sparse regularized version of the sample covariance of these values. If this approach is taken, it is usually feasible to generate more than N such values in order to reduce the Monte Carlo error in the regularized sample covariance matrix.

4·3. *Covariance inflation*

A common tuning heuristic to combat filter divergence due to, e.g., Monte Carlo errors or model bias is covariance inflation (Anderson, 2007). This means that the forecast ensemble $\{x_j^p\}$ is artificially inflated before the update is performed, $x_j^p \mapsto \bar{x}^p + \delta(x_j^p - \bar{x}^p)$, where $\bar{x}^p = N^{-1}\sum_{j=1}^N x_j^p$ is the sample mean and $\delta \geqslant 1$ some inflation factor. In principle, covariance inflation can be used in conjunction with any Monte Carlo filter that operates on the forecast ensemble, i.e., also with the ensemble Kalman particle filter.

5. CHOICE OF γ

5·1. *Asymptotics of* π^u_{EnKPF}

It is not difficult to see that if $\{x_j^p\}$ is an independent and identically distributed sample from π^p with finite second moments, then for any $\gamma \in [0, 1]$ and any fixed observation y, the random probability measure π^u_{EnKPF} as defined in (6) converges almost surely weakly to a nonrandom limit distribution $\pi^{u,\infty}_{\mathrm{EnKPF}}$ as $N \to \infty$. If π^p is Gaussian, the limit distribution equals the true posterior $\pi^u(\mathrm{d}x) \propto \varphi(y; Hx, R)\pi^p(\mathrm{d}x)$. If π^p is non-Gaussian, the limit distribution depends on γ and generally differs from the correct posterior if $\gamma > 0$. Unfortunately, the limit cannot be easily identified, and in particular it is difficult to quantify the systematic error as a function of γ. Using arguments similar to Randles (1982), it is also possible to show that for bounded test functions h,

$$N^{1/2}\left\{\int h(x)\pi^u_{\mathrm{EnKPF}}\,\mathrm{d}x - \int h(x)\pi^{u,\infty}_{\mathrm{EnKPF}}\,\mathrm{d}x\right\} \to \mathcal{N}(0, V)$$

weakly, where the asymptotic covariance V depends on h, π^p and γ. In general, V is analytically intractable, and for instance, we cannot verify whether V decreases as a function of γ, as one would expect. The bottom line is that the asymptotics of π^u_{EnKPF} do not lead to practically exploitable rules on how to choose γ by balancing bias and variance.

5·2. *Asymptotics of weights*

Recall that for $\gamma = 0$ the method is exactly the particle filter, and for $\gamma = 1$ it is exactly the ensemble Kalman filter. Hence it is clear that there is a range of values γ where we obtain an interesting compromise between the two methods in the sense that the weights $\{\alpha_i^{u,\gamma}\}$ are neither uniform nor degenerate. We try to provide some theoretical insight where this range of values γ is, and later we develop a criterion that chooses a good value γ automatically.

We want to see how the weights $\{\alpha_i^{u,\gamma}\}$ in (6) behave as a function of γ when the dimension of the observations is large. By definition

$$\begin{aligned}\alpha_i^{u,\gamma} &\propto \exp\left\{-\frac{1}{2}(y - Hv_i^{u,\gamma})'\left\{HQ(\gamma, \hat{P}^p)H' + \frac{1}{1-\gamma}R\right\}^{-1}(y - Hv_i^{u,\gamma})\right\}\\ &\propto \exp\left\{-\frac{1}{2}(x_i^p - \mu^p)'\hat{C}_\gamma(x_i^p - \mu^p) + \hat{d}_\gamma'(x_i^p - \mu^p)\right\}\end{aligned}$$

where μ^p is the prediction mean,

$$\begin{aligned}\hat{C}_\gamma &= (1-\gamma)H'(I - \hat{K}_\gamma' H')\{(1-\gamma)H\hat{Q}_\gamma H' + R\}^{-1}(I - H\hat{K}_\gamma)H,\\ \hat{d}_\gamma &= (1-\gamma)H'(I - \hat{K}_\gamma' H')\{(1-\gamma)H\hat{Q}_\gamma H' + R\}^{-1}(I - H\hat{K}_\gamma)(y - H\mu^p)\end{aligned}$$

and $\hat{K}_\gamma$ and $\hat{Q}_\gamma$ stand for $K(\gamma\hat{P}^p)$ and $Q(\gamma, \hat{P}^p)$.

The following lemma gives an approximate formula for the variance of $\alpha_i^{u,\gamma}$.

LEMMA 1. *Define approximate weights by*

$$\tilde{\alpha}_i^{u,\gamma} = \frac{1}{N}\frac{\exp\{-\frac{1}{2}(x_i^p - \mu^p)'C_\gamma(x_i^p - \mu^p) + d_\gamma'(x_i^p - \mu^p)\}}{E[\exp\{-\frac{1}{2}(x_i^p - \mu^p)'C_\gamma(x_i^p - \mu^p) + d_\gamma'(x_i^p - \mu^p)\}]}.$$

where C_γ and d_γ are as defined above, but with the true forecast covariance P^p instead of $\hat{P}^p$. If the forecast sample is independent and identically $\mathcal{N}(\mu^p, P^p)$ distributed, then

$$N^2\mathrm{var}(\tilde{\alpha}_i^{u,\gamma}) = \frac{\det(P^pC_\gamma + I)}{\det(2P^pC_\gamma + I)^{1/2}}\exp(d_\gamma'[\{C_\gamma + (P^p)^{-1}/2\}^{-1} - \{C_\gamma + (P^p)^{-1}\}^{-1}]d_\gamma) - 1.$$

Moreover,

$$\lim_{\gamma\uparrow 1}\frac{N^2\mathrm{var}(\tilde{\alpha}_i^{u,\gamma})}{(1-\gamma)^2} = \frac{1}{2}\mathrm{tr}(HP^pH'M) + (y - H\mu^p)'M(y - H\mu^p), \tag{8}$$

where $M = (I - K_1'H')R^{-1}(I - HK_1)HP^pH'(I - K_1'H')R^{-1}(I - HK_1)$.

A proof is given in the Appendix. The matrix M is positive definite and also the trace in formula (8) is positive. Therefore we expect that the variance of the weights is of the order

$\mathcal{O}\{N^{-2}(1-\gamma)^2 r\}$, where r is the dimension of the observations. This is true if P^p, R and H are all multiples of the identity, and there is no reason for different behaviour in other cases. This suggests that for high-dimensional observations, we need to choose γ close to one in order to avoid degeneracy. For any fixed γ, the weights will collapse as $r \to \infty$, and the asymptotic behaviour is the same as for the particle filter examined in Snyder et al. (2008). The final update can still differ from the ensemble Kalman filter update, even if its largest part occurs with the ensemble Kalman filter.

5·3. *Criteria for the selection of γ*

We aim to choose the smallest value of γ that still gives a realistic amount of uncertainty about the value of the state. If we believe that the uncertainty contained in the ensemble Kalman filter is reasonable and if the spread of the sample is considered to be a meaningful measure of uncertainty, we can choose γ such that the spread of the update is not smaller than a factor τ times the spread of an ensemble Kalman filter update, where τ is maybe between 0·5 and 0·8. More precisely, denoting the standard deviations of the Gaussian mixture (6) by σ_k^u and the square roots of the diagonal elements of $\{I - K(\hat{P}^p)H\}\hat{P}^p$ by $\sigma_k^{u,En}$, we could take the smallest γ such that $\sum_k \sigma_k^u \geqslant \tau \sum_k \sigma_k^{u,En}$. If we want to control not only the total spread but all marginal spreads, we would take γ such that $\sum_k \min\{1, \sigma_k^u (\sigma_k^{u,En})^{-1}\} \geqslant \tau q$.

However, there are two problems with this approach. First, computing σ_k^u is demanding because we have to compute among other things $P^{u,\gamma}$; cf. the discussion in § 4. Estimating σ_k^u and $\sigma_k^{u,En}$ from the final update samples is simpler, but this would then depend also on the generated noise variables. Although this can be reduced somewhat by taking the same realizations of $\epsilon_{1,j}$ and $\epsilon_{2,j}$ for all values of γ under consideration, it is not satisfactory.

A second argument against a choice of γ based on the spread of the update distribution is that it is not necessarily a monotone function of γ because the variance of the correct posterior is only on average smaller than the variance $\{I - K(P^p)H\}P^p$ of the ensemble Kalman filter update. If the prior is multimodal, the correct update can have a larger variance for some values of y.

For these reasons we decided to look only at the weights $\alpha_i^{u,\gamma}$ and to measure the sampling diversity by the so-called effective sample size (Liu, 1996), which is defined as

$$\text{ESS} = \frac{1}{\sum_i (\alpha_i^{u,\gamma})^2} = \frac{N}{1 + N \sum_i (\alpha_i^{u,\gamma} - 1/N)^2}.$$

We then choose γ as the smallest value for which $\text{ESS} > \tau N$. This criterion does not take into account the spread in $P^{u,\gamma}$, which also increases if γ increases. Therefore, it gives only a lower bound for the diversity, but it is easy to compute. In order to avoid an extensive search over $\gamma \in [0, 1]$, in the examples below we considered only multiples of $1/15$ as values for γ and used a binary search tree that results in at most four search steps, assuming that the diversity is increasing in γ. We did not try to prove this assumption since the calculation is expected to be extremely tedious; the assumption is safe to make, though, since at the worst we end up with γ too large. Alternatively one could use an approximation of ESS based on (8).

6. EXAMPLES OF SINGLE UPDATES

6·1. *Description of the set-up*

We consider a single update for four situations. In all cases $H = I$ and $R = \sigma^2 I$. There are two forecast samples $\{x_j^p\}$ combined with two values $y = y_1$ and $y = y_2$ for the observations. The

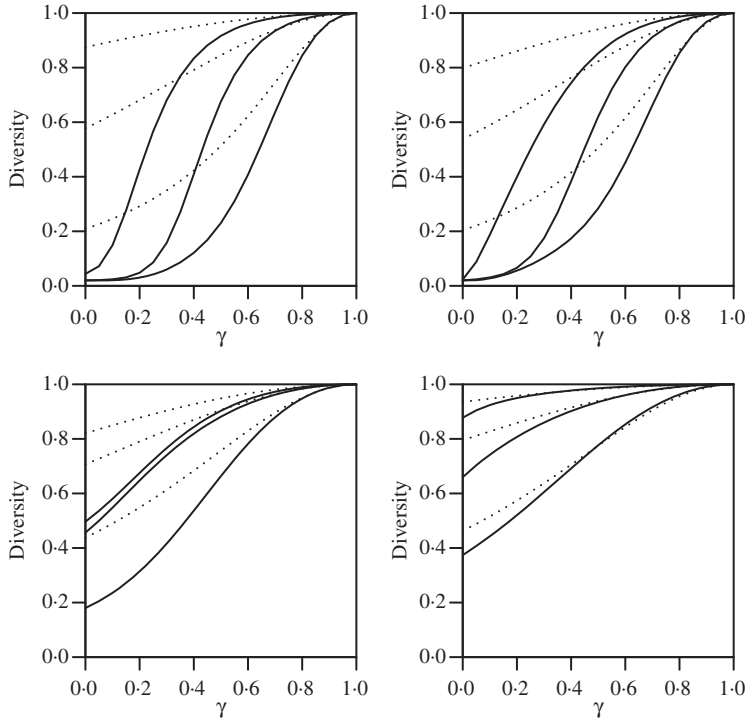

Fig. 1. Diversity $N^{-1} \times \text{ESS}$ as a function of γ. Top row: Gaussian prior, bottom row: bimodal prior. Left column: y_1, right column: y_2. Dimension q of the state variable (from top line to bottom line): $q = 10, 50, 250$. The solid lines show $N^{-1} \times \text{ESS}$, whereas the dotted lines show the approximations to this quantity computed from (8) with μ^p and P^p estimated from the sample.

construction of the forecast sample starts with a sample $z_j \sim \mathcal{N}_q(0, I)$, which is then modified to introduce non-Gaussian features. More precisely, we consider the following situations:

1. A Gaussian prior: we set $\sigma = 0{\cdot}5$, $x_j^p = z_j$, $y_1 = (0, \ldots, 0)'$, and $y_2 = (1{\cdot}5, 1{\cdot}5, 0, \ldots, 0)'$. This means that the second observation contradicts the prior, although not excessively.
2. A bimodal prior: we set $\sigma = 3$, $x_j^p = z_j$ for $j \leqslant N/2$ and $x_j^p = z_j + (6, 0, \ldots, 0)'$ for $j > N/2$, $y_1 = (-2, 0, \ldots, 0)'$ and $y_2 = (3, 0, \ldots, 0)$. In the former case, the true posterior is unimodal and in the latter case it is bimodal.

We take $N = 50$ and $q = 10, 50, 250$. We generate one sample in dimension 250 and use the first q components. In all cases, we use a triangular taper with range 10, assuming that the states are values along a line; the optimal taper has range 0, since the true covariance P^p is diagonal. Without a taper, all procedures including the ensemble Kalman filter break down: the spread becomes much too low as the dimension increases, and the weights of the ensemble Kalman particle filter become more balanced as the dimension increases.

6·2. *Variation of γ*

We compute the update for $\gamma \in \{0, 0{\cdot}05, 0{\cdot}10, \ldots, 1\}$ and take as the measure of the sampling diversity the quantity $N^{-1} \times \text{ESS}$ introduced above. Some results are shown in Fig. 1.

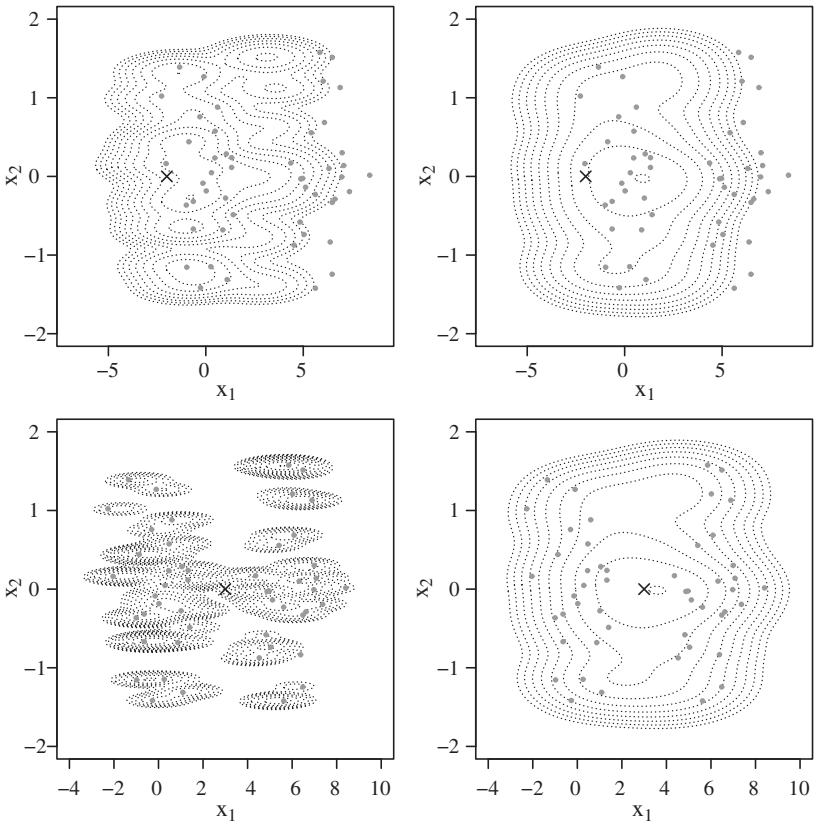

Fig. 2. First two coordinates of the update of the bimodal prior with $q = 250$. Upper row: y_1, lower row: y_2. Left column: ensemble Kalman particle filter with γ chosen to achieve a diversity of about 40%, which corresponds to $\gamma = 0{\cdot}3$ for the upper plot, and $\gamma = 0{\cdot}05$ for the lower plot. Right column: ensemble Kalman filter. The prior sample is given by the light grey dots, the observation is marked with a cross, and the contours show the Gaussian mixture (6): levels are equidistant on a log scale such that the lowest level corresponds to 1% of the maximum.

The diversity increases with γ and decreases with the dimension q as expected. In the bimodal case shown here, even the particle filter does not degenerate, and small or moderate values of γ apparently give sufficient diversity. Of course, the diversity for $\gamma = 0$, i.e., for the particle filter, depends not only on the shape of the prior but also on other specifications of the problem, and no simple general rule can be deduced.

6·3. *Updates of the first two coordinates*

We concentrate on the first two coordinates of $x_j^{u,\gamma}$ that contain the non-Gaussian features, if present. We show the contours of the true update density (6) for the ensemble Kalman filter and for the filter with γ chosen such that the diversity $\tau = N^{-1} \times \mathrm{ESS}$ is approximately 40%. Figure 2 shows the results for the bimodal prior with $q = 250$. In case of the Gaussian prior, the two plots, which are not shown here, are virtually identical. In the non-Gaussian situation, the combined filter is able to pick up some non-Gaussian features. In particular, the shape and not only the location depends on the observation, and the bimodality of the posterior is captured.

7. Examples of filtering with many cycles

7·1. *The Lorenz 96 model*

The 40-variable configuration of the Lorenz 96 model (Lorenz & Emanuel, 1998) is governed by the ordinary differential equation

$$\frac{\mathrm{d}X_t^k}{\mathrm{d}t} = (X_t^{k+1} - X_t^{k-2})X_t^{k-1} - X_t^k + 8 \quad (k = 1, \ldots, 40)$$

where the boundary conditions are assumed to be cyclic, i.e., $X^k = X^{40+k}$. The model is chaotic and mimics the time evolution of a scalar meteorological quantity on a latitude circle. We adopt the experimental set-up of Bengtsson et al. (2003), Lei & Bickel (2011) and Frei & Künsch (2013): measurements of odd components X^{2k-1} with uncorrelated additive $\mathcal{N}(0, 0{\cdot}5)$ noise at observation times $0{\cdot}4 \times n$ $(n = 1, \ldots, 2000)$ are taken. The large lead time produces a strongly nonlinear propagation step. The system is integrated using Euler's method with step size 0·001. Both the ensemble Kalman filter and ensemble Kalman particle filter are run with $N = 400$ ensemble members. The true initial state and the initial ensemble members are randomly drawn from $\mathcal{N}_{40}(0, I)$. All sample covariance matrices are replaced by tapered estimates; for the sake of simplicity, we used the same taper matrix C throughout, with

$$C_{kl} = \mathcal{C}_0\{\min(|k - l|, 40 - |k - l|)\},$$

where $\mathcal{C}_0$ is the correlation function given in Gaspari & Cohn (1999, equation (4.10)) with support half-length $c = 10$. No covariance inflation is used, in contrast to Lei & Bickel (2011) where the inflation factors are presumably determined by off-line tuning. For the ensemble Kalman particle filter, the parameter γ is chosen adaptively to ensure that the diversity $\tau = N^{-1} \times \mathrm{ESS}$ stays within a prespecified interval $[\tau_0, \tau_1] \subset [0, 1]$ if possible; the resampling uses balanced sampling.

The filter performance is assessed via a scoring rule evaluated at observation times. Here, we use the root mean square error of the ensemble mean, and the continuous ranked probability score (Gneiting et al., 2007) for the first two state variables. More precisely, if X_t^k is the true solution at time t, and $\hat{X}_t^k$ the mean of the updated ensemble, and $\hat{F}_t^k(y)$ the marginal empirical cumulative distribution function of the updated ensemble, then the root mean square error of the ensemble mean at time t is

$$\mathrm{RMSE}_t = \left\{ \frac{1}{q} \sum_{k=1}^{q} (X_t^k - \hat{X}_t^k)^2 \right\}^{1/2} \tag{9}$$

and the continuous ranked probability score for the kth variable at time t is

$$\mathrm{CRPS}_t^k = \int_{\mathbb{R}} \left\{ \hat{F}_t^k(y) - 1_{\{y \geqslant X_t^k\}} \right\}^2 \mathrm{d}y \quad (k = 1, 2) \tag{10}$$

where $t = 0{\cdot}4 \times n$ $(n = 1, \ldots, 2000)$. For reasons of symmetry, we consider only the continuous ranked probability score of the first two state variables, one observed and one unobserved. Tables 1 and 2 compile the first and ninth deciles, mean and median of the scores. The gain over the ensemble Kalman filter achieved by the ensemble Kalman particle filter is substantial. In particular, as the continuous ranked probability scores show, the ensemble Kalman particle filter is able to track the unobserved states much more accurately than is the ensemble Kalman filter. Overall, the relative improvement over the ensemble Kalman filter is comparable with the improvement reported for the localized nonlinear ensemble adjustment filter with first-order correction in Lei & Bickel (2011) and for the mixture ensemble Kalman filter in Frei & Künsch

Table 1. *Summary statistics of* RMSE (9) *over* 2000 *cycles for the Lorenz* 96 *system, experimental set-up as given in* § 7·1

$N = 400$	$[\tau_0, \tau_1]$	10%	50%	Mean	90%
EnKF		0·56	0·81	0·87	1·25
EnKPF	[0·80, 0·90]	0·52	0·75	0·83	1·21
EnKPF	[0·50, 0·80]	0·51	0·73	0·80	1·18
EnKPF	[0·30, 0·60]	0·50	0·71	0·79	1·17
EnKPF	[0·25, 0·50]	0·49	0·70	0·78	1·16
EnKPF	[0·10, 0·30]	0·49	0·71	0·79	1·17

Ensemble Kalman filter (EnKF) and ensemble Kalman particle filter (EnKPF) with constrained diversity $\tau = N^{-1} \times \text{ESS} \in [\tau_0, \tau_1]$ for the weights.

Table 2. *Summary statistics of* CRPS (10) *over* 2000 *cycles for the state variables* X^1 (*observed*) *and* X^2 (*unobserved*) *of the Lorenz* 96 *system, experimental set-up as given in* § 7·1

		X^1				X^2			
$N = 400$	$[\tau_0, \tau_1]$	10%	50%	Mean	90%	10%	50%	Mean	90%
EnKF		0·12	0·22	0·32	0·65	0·14	0·38	0·57	1·18
EnKPF	[0·80, 0·90]	0·11	0·21	0·30	0·62	0·13	0·33	0·54	1·13
EnKPF	[0·50, 0·80]	0·11	0·21	0·29	0·61	0·12	0·32	0·51	1·10
EnKPF	[0·30, 0·60]	0·11	0·20	0·29	0·59	0·12	0·32	0·49	1·02
EnKPF	[0·25, 0·50]	0·10	0·20	0·28	0·58	0·11	0·31	0·48	1·00
EnKPF	[0·10, 0·30]	0·10	0·21	0·29	0·59	0·11	0·31	0·50	1·05

Ensemble Kalman filter (EnKF) and ensemble Kalman particle filter (EnKPF) with constrained diversity $\tau = N^{-1} \times \text{ESS} \in [\tau_0, \tau_1]$ for the weights.

(2013). Arguably, the best performance of the ensemble Kalman particle filter is achieved with diversity constrained to [0·25, 0·50], but the scores are surprisingly robust.

To shed some light on our heuristic to adaptively choose γ, Fig. 3 shows a time series plot of these values for one of the runs; the following remarks also apply to the other runs. The series appears to be stationary, and the running mean of the values quickly stabilizes. We also re-ran the filter with γ fixed to the mean, which did not noteworthily impact the filter performance. Practically speaking, one could in fact monitor the sequence for convergence using, say, standard Markov chain Monte Carlo diagnostics, and after the spin-up, work with the mean as static value for γ to reduce the computational load.

For smaller ensemble sizes, the ensemble Kalman particle filter can still improve on the ensemble Kalman filter, but the results are less impressive. Tables 3 and 4 compile the summaries. Obviously, the range of constraints for the diversity that lead to an improvement over the ensemble Kalman filter narrows when the ensemble size decreases. For the nonlinear ensemble adjustment filter from Lei & Bickel (2011) we were not able to obtain a stable run with $N = 100$ particles even with extensive trial and error for the tunable parameter, that is, the window sizes for the sliding-window localization and the inflation amount. The same applies for the filter proposed in Papadakis et al. (2010). Moreover, for a particle filter, many more particles are required to compete with the ensemble Kalman filter, as illustrated in Bocquet et al. (2010) for a slightly

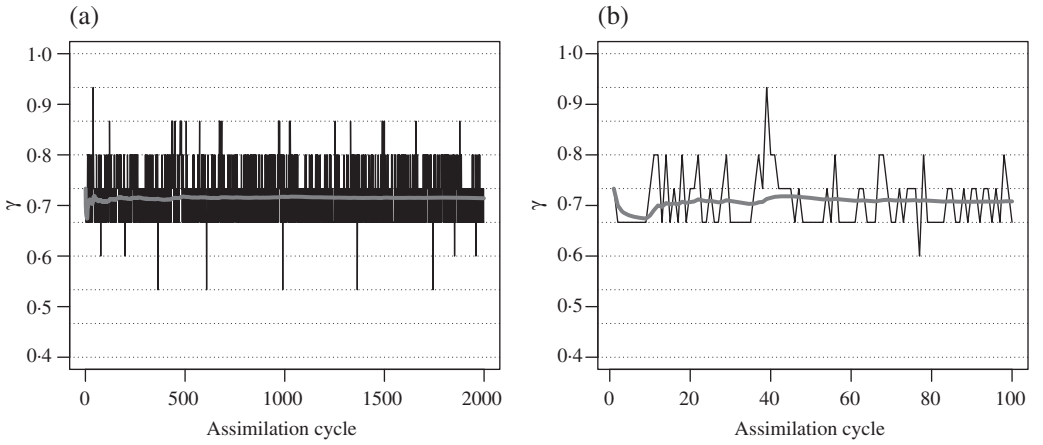

Fig. 3. Lorenz 96 system, experimental set-up as given in § 7·1: time series plot (black line) of the adaptively chosen values of γ with diversity τ constrained to [0·80, 0·90]. (a) shows the whole 2000 cycles, whereas (b) gives a zoomed-in view of the first 100 cycles. The running mean is shown in grey.

Table 3. *As Table 1, but for* $N = 100$ *particles*

$N = 100$	$[\tau_0, \tau_1]$	10%	50%	Mean	90%
EnKF		0·57	0·82	0·94	1·41
EnKPF	[0·80, 0·90]	0·55	0·81	0·94	1·48
EnKPF	[0·50, 0·80]	0·54	0·81	0·95	1·52
EnKPF	[0·30, 0·60]	0·55	0·82	0·97	1·57
EnKPF	[0·25, 0·50]	0·54	0·84	1·00	1·75
EnKPF	[0·10, 0·30]	0·69	1·37	1·61	2·93

Table 4. *As Table 2, but for* $N = 100$ *particles*

		X^1				X^2			
$N = 100$	$[\tau_0, \tau_1]$	10%	50%	Mean	90%	10%	50%	Mean	90%
EnKF		0·12	0·23	0·34	0·69	0·14	0·38	0·68	1·47
EnKPF	[0·80, 0·90]	0·11	0·23	0·34	0·70	0·13	0·37	0·66	1·48
EnKPF	[0·50, 0·80]	0·11	0·23	0·33	0·68	0·13	0·36	0·70	1·46
EnKPF	[0·30, 0·60]	0·11	0·22	0·34	0·71	0·13	0·36	0·66	1·44
EnKPF	[0·25, 0·50]	0·10	0·23	0·33	0·68	0·12	0·36	0·68	1·52
EnKPF	[0·10, 0·30]	0·11	0·30	0·52	1·14	0·14	0·55	1·18	3·02

different configuration of the stochastic Lorenz model. We also did some tests with fewer particles, but things become very unstable, and even the ensemble Kalman filter with carefully tuned covariance inflation tends to diverge for $N = 50$ particles.

7·2. *A many-species Lotka–Volterra model*

We consider the following spatiotemporal chaotic Lotka–Volterra model introduced and studied in detail in Sprott et al. (2005):

$$\frac{\mathrm{d}Z_t^k}{\mathrm{d}t} = Z_t^k(1 - Z_t^{k-2} - Z_t^k - Z_t^{k+1}) \quad (k = 1, \ldots, 100)$$

Table 5. *Summary statistics of* $10 \times$ RMSE (9) *over* 2000 *cycles for the Lotka–Volterra model, experimental set-up as given in* § 7·2

$N = 50$	$[\tau_0, \tau_1]$	10%	50%	Mean	90%
EnKF		1·28	1·95	2·01	2·75
EnKPF	[0·80, 0·90]	1·17	1·86	1·93	2·69
EnKPF	[0·50, 0·80]	1·06	1·73	1·78	2·55
EnKPF	[0·30, 0·60]	1·00	1·71	1·86	2·73
EnKPF	[0·25, 0·50]	1·03	1·77	1·94	2·92
EnKPF	[0·10, 0·30]	1·25	2·49	3·40	6·72

Ensemble Kalman filter (EnKF) and ensemble Kalman particle filter (EnKPF) with constrained diversity $\tau = N^{-1} \times \text{ESS} \in [\tau_0, \tau_1]$ for the weights.

with cyclic boundary conditions, $Z^k = Z^{100+k}$. This system describes 100 populations of identical species located at 100 different sites along a circle. Each population competes for resources with two of its four nearest neighbours, and $Z_t^k \in (0, 1)$ is the size at time t of the kth population measured in units of the carrier capacity relative to no competition. Since the solutions are constrained to the unit cube, we consider the system on the logistic scale

$$X_t^k = \text{logit}(Z_t^k) = \log \frac{Z_t^k}{1 - Z_t^k}.$$

This transformation is natural since the decoupled system $\mathrm{d}\widetilde{Z}_t^k/\mathrm{d}t = \widetilde{Z}_t^k(1 - \widetilde{Z}_t^k)$, which is commonly known as the logistic equation, satisfies $\mathrm{d}\widetilde{X}_t^k/\mathrm{d}t = 1$.

We adopt a similar experimental design as for the Lorenz 96 model: measurements of odd components X^{2k-1} with uncorrelated additive $\mathcal{N}(0, 0{\cdot}5)$ noise at observation times $50 \times n$ ($n = 1, \ldots, 2000$) are taken. The large lead time leads to a strongly nonlinear propagation step (Sprott et al., 2005, § 3). The system is integrated using the common explicit fourth-order Runge–Kutta method with step size 0·05. The true initial state and the initial ensemble members are randomly drawn from $\mathcal{N}_{100}(m, I)$, where $m^k = \text{logit}(1/3)$ ($k = 1, \ldots, 100$) is an unstable equilibrium point that corresponds to coexistence of the populations.

Both the ensemble Kalman filter and ensemble Kalman particle filter, with adaptively chosen γ, are run with $N = 50$ ensemble members, i.e., $N < q$. The Gaspari–Cohn taper as described in § 7·1 is used with $c = 25$, no covariance inflation is applied. The scores are compiled in Tables 5 and 6. The gain over the ensemble Kalman filter achieved by the ensemble Kalman particle filter is substantial. We were not able to obtain a stable run for the localized nonlinear ensemble adjustment filter (Lei & Bickel, 2011). With minor tuning, the mixture ensemble Kalman filter (Frei & Künsch, 2013) runs stably, but produces worse results than the ensemble Kalman filter.

7·3. *The Korteweg–de Vries equation*

We consider the Korteweg–de Vries equation on the circle (Drazin & Johnson, 1989):

$$\partial_t X + \partial_s^3 X + 3\partial_s X^2 = 0$$

with domain $(s, t) \in [-1, 1) \times [0, \infty)$ and periodic boundary conditions, $X(s = -1, t) = X(s = 1, t)$. Variants of this equation have been used as test beds for data assimilation in, e.g., van Leeuwen (2003), Lawson & Hansen (2005), or Zupanski & Zupanski (2006). The spatial

Table 6. *Summary statistics of* 10 × CRPS (10) *over* 2000 *cycles for the state variables* X^1 (*observed*) *and* X^2 (*unobserved*) *of the Lotka–Volterra model, experimental set-up as given in* § 7·2

		X^1				X^2			
$N = 50$	$[\tau_0, \tau_1]$	10%	50%	Mean	90%	10%	50%	Mean	90%
EnKF		0·15	0·63	0·98	2·26	0·16	0·67	1·03	2·44
EnKPF	[0·80, 0·90]	0·13	0·61	0·95	2·16	0·13	0·60	0·96	2·19
EnKPF	[0·50, 0·80]	0·11	0·53	0·87	2·08	0·11	0·57	0·87	2·03
EnKPF	[0·30, 0·60]	0·10	0·55	0·90	2·08	0·11	0·55	0·92	2·10
EnKPF	[0·25, 0·50]	0·10	0·55	0·94	2·22	0·10	0·58	0·99	2·43
EnKPF	[0·10, 0·30]	0·12	0·77	1·83	4·26	0·13	0·80	2·00	4·76

Ensemble Kalman filter (EnKF) and ensemble Kalman particle filter (EnKPF) with constrained diversity $\tau = N^{-1} \times \text{ESS} \in [\tau_0, \tau_1]$ for the weights.

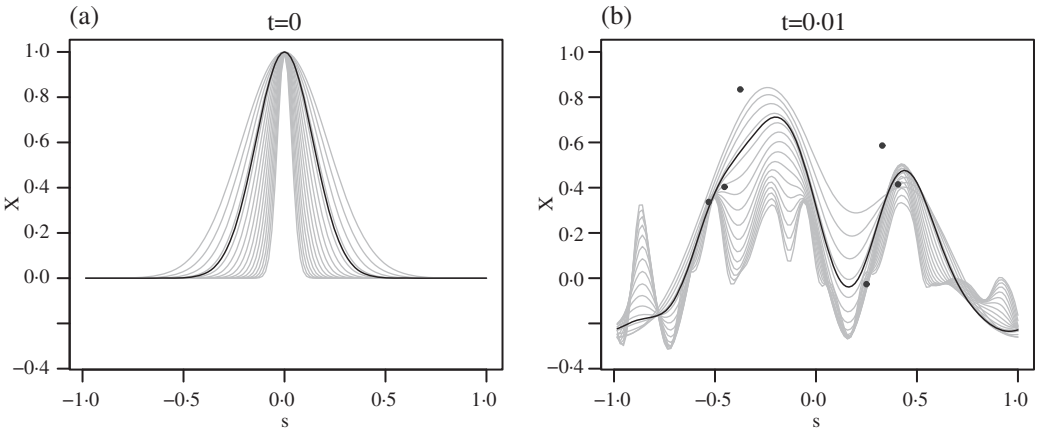

Fig. 4. Korteweg–de Vries equation, experimental set-up as given in § 7·3. (a) shows the initial 16-member ensemble (grey) at time $t = 0$ together with the true solution (black). (b) shows the forecast ensemble (grey) at the first observation time $t = 0{\cdot}01$ together with the observations (black bullets) and the true solution (black).

domain $[-1, 1)$ is discretized using an equispaced grid with $q = 128$ grid points. The spectral split step method is used to solve the equation numerically, with an explicit fourth-order Runge–Kutta time-step with step size 10^{-4} for the nonlinear part of the equation. As prior we take the random field

$$X(s, t = 0) = \exp\left(-\frac{s^2}{\eta^2}\right), \quad \log(\eta) \sim \mathcal{U}\{\log(0{\cdot}05), \log(0{\cdot}3)\}.$$

For the truth, we use $\eta = 0{\cdot}2$. The initial ensemble is a quasi-random sample from $X(\cdot, t = 0)$. The ensemble size is $N = 16$, and thus $N \ll q$. Six irregularly spaced observations with uncorrelated additive $\mathcal{N}(0, 0{\cdot}02)$ noise at observation times $0{\cdot}01 \times n$ $(n = 1, \ldots, 10)$ are taken. The lead time is rather short and corresponds to weakly nonlinear behaviour. For illustration, Fig. 4 displays the initial 16-member ensemble and the forecast ensemble at the first observation time together with the observations.

The particle filter, the ensemble Kalman filter and the ensemble Kalman particle filter are run without tapering, for reasons discussed below. For the ensemble Kalman particle filter, we fix $\gamma = 0{\cdot}05$, which ensures that $\tau = N^{-1} \times \text{ESS}$ lies roughly in the interval $[0{\cdot}80, 0{\cdot}90]$. Since the particle filter degenerates very quickly for such a small ensemble, a benchmark run with $N = 256$

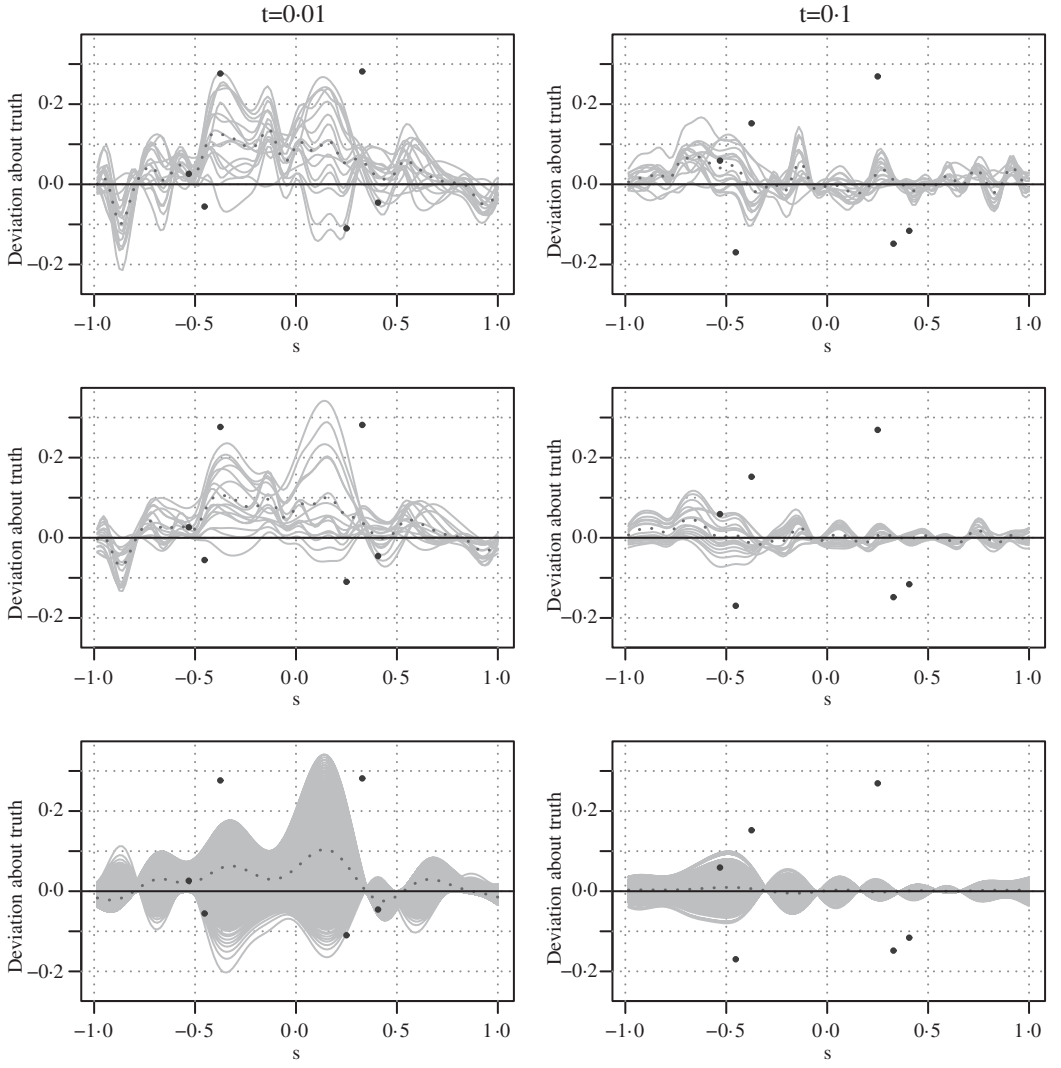

Fig. 5. Korteweg–de Vries equation, experimental set-up as given in § 7·3. Ensemble deviations about the truth, i.e., filtered ensemble minus true solution, after the 1st ($t = 0{\cdot}01$, left panel) and 10th ($t = 0{\cdot}1$, right panel) update cycle, for the ensemble Kalman filter and ensemble Kalman particle filter with $N = 16$ particles (top two rows), and for the particle filter with $N = 256$ particles (bottom row). The solid grey lines are the deviations, the dotted grey lines the average of the deviations, and the black bullets are the observations minus the truth.

particles is carried out using balanced resampling as in Künsch (2005, equation (12)). Figure 5 displays ensemble deviations from the true solution after 1 and 10 update cycles. Apparently, both the ensemble Kalman filter and ensemble Kalman particle filter track the true solution reasonably well, and the state uncertainty is well represented. In terms of error of the ensemble mean, there is not much difference between the ensemble Kalman filter and the ensemble Kalman particle filter. However, the ensemble Kalman particle filter produces particles that exhibit fewer dynamical inconsistencies. In Fig. 6, for each filter, the particle with the most curvature after 10 update cycles is shown, where the curvature of a solution $X(s, t)$ is defined by $\int_{-1}^{1} |\partial_s^2 X| \{1 + (\partial_s X)^2\}^{-3/2} \, \mathrm{d}s$, and a finite difference approximation is used for the discretized solutions. The true solution, which is not shown in the plots, is virtually identical to the particle shown in the

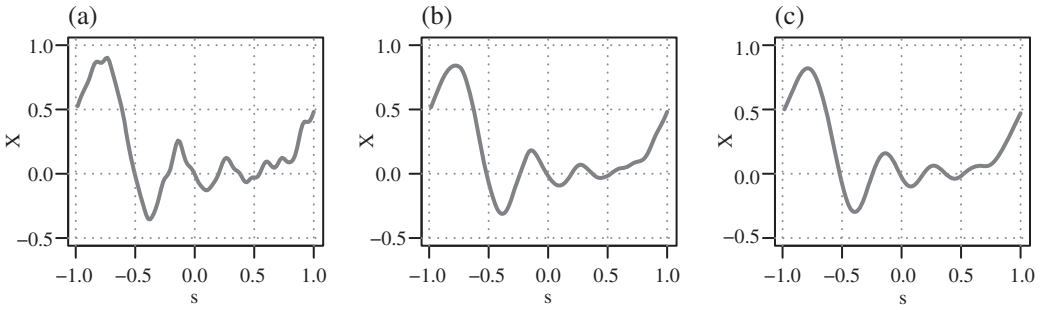

Fig. 6. Korteweg–de Vries equation, experimental set-up as given in § 7·3. Particle with most curvature after 10 update cycles ($t = 0{\cdot}1$), for the ensemble Kalman filter (a), ensemble Kalman particle filter (b) and particle filter (c).

rightmost plot. The particle filter, which conserves dynamical constraints, yields very smooth particles, whereas the ensemble Kalman filter may produce wiggly, unphysical particles. The ensemble Kalman particle filter lies in between them: some particles still show slight dynamical imbalances, but these are much less pronounced than for the ensemble Kalman filter. Also, if a taper is applied to the forecast covariance matrix, which is not the case in the example here, the ensemble Kalman filter suffers even more from these imbalances.

8. Possible generalizations

In the spirit of the progressive correction idea (Musso et al., 2001), the ensemble Kalman particle filter update could also be split up in several steps. We fix constants $\gamma_i > 0$ and $\delta_i > 0$ with $\sum_{i=1}^{L} \gamma_i + \delta_i = 1$. Then, for $i = 1, \ldots, L$, we apply an ensemble Kalman filter update with likelihood $\ell(x \mid y)^{\gamma_i}$, followed by a particle filter update with likelihood $\ell(x \mid y)^{\delta_i}$, followed by the resampling step. It is necessary to estimate the predictive covariance only for the first step; for the subsequent steps, $i = 2, \ldots, L$, we can compute the covariance analytically from the mixture representation (6); for large q, this is numerically delicate, but the remedies discussed in § 4 can be applied. We expect that the bias of such an iterative ensemble Kalman particle filter update is similar as for a single ensemble Kalman particle filter update with $\gamma = \sum_{i=1}^{L} \gamma_i$, but the variance will decrease with increasing L since the likelihoods become flatter. In the limiting case $\sum_{i=1}^{L} \gamma_i \to 0$, which corresponds to a full particle filter update, we conjecture that $L = \mathcal{O}(r)$ is sufficient to retain the sampling diversity. This claim is supported by recent work of Alexandros Beskos et al., from University College London who analyse the tempering idea in simpler but related situations.

A potential drawback of the ensemble Kalman particle filter in comparison to non-Gaussian ensemble filters akin to Lei & Bickel (2011) is its restriction to Gaussian linear observations. However, the idea of combining an ensemble Kalman filter and a particle filter update could also be used for arbitrary observation densities. Let H be a matrix that selects those components of the state variable that influence the observation, and assume that we have an approximation of the likelihood of the form $\ell(Hx \mid y) \approx \varphi\{g(y); Hx, R(y)\}$. Then we can use this approximation for an ensemble Kalman filter update, and correct by a particle filter update with weights proportional to $\ell(Hx_j^u \mid y)[\varphi\{g(y); Hx_j^u, R(y)\}]^{-1}$. In order to construct an approximation of the likelihood of the above form, we can use a Taylor approximation

$$\log \ell(Hx \mid y) \approx \log \ell(H\mu^p \mid y) + a(y)'H(x - \mu^p) + \frac{1}{2}(x - \mu^p)'H'b(y)H(x - \mu^p)$$

where $a(y)$ and $b(y)$ are the gradient and the Hessian, respectively, of the loglikelihood. Then $R(y) = -b(y)^{-1}$ and $g(y) = R(y)a(y)$. Alternatively, one could centre the expansion at the mode of the likelihood. Such an approximation is expected to work well in cases where the likelihood is log-concave, e.g., when y given x is Poisson with parameter $\exp(x)$. An obvious drawback to this approach is that the resampling step produces tied values, which makes the method unsuitable for deterministic models. A simple remedy is to use kernel resampling with shrinkage (Liu & West, 2001). We tested this for the set-up described in Lei & Bickel (2011, § 4.b.2), where the observation errors are generated from a Laplace distribution. For $N = 400$ particles, the relative improvement over the ensemble Kalman filter is competitive with the improvement reported for the nonlinear ensemble adjustment filter, but for smaller ensemble sizes, our method gives poorer results than the ensemble Kalman filter.

ACKNOWLEDGEMENT

The authors thank Jo Eidsvik for fruitful discussions, and an associate editor and two referees for helpful comments.

APPENDIX

Proof of Lemma 1

We set

$$Z = \exp\left\{-\frac{1}{2}(x_i^p - \mu^p)' C_\gamma (x_i^p - \mu^p) + d_\gamma'(x_i^p - \mu^p)\right\}.$$

Then by definition

$$\mathrm{var}(\tilde{\alpha}_i^{u,\gamma}) = \frac{1}{N^2}\left\{\frac{E(Z^2)}{E(Z)^2} - 1\right\}.$$

The lemma follows by completing the square and recalling that Gaussian densities integrate to one. More precisely, for any x and any positive definite matrix Γ:

$$x'(C_\gamma + \Gamma)x - 2d_\gamma' x = \{x - (C_\gamma + \Gamma)^{-1} d_\gamma\}'(C_\gamma + \Gamma)\{x - (C_\gamma + \Gamma)^{-1} d_\gamma\} - d_\gamma'(C_\gamma + \Gamma)^{-1} d_\gamma.$$

Therefore

$$E(Z) = [\det(P^p)\det\{C_\gamma + (P^p)^{-1}\}]^{-1/2} \exp\left[\frac{1}{2} d_\gamma' \{C_\gamma + (P^p)^{-1}\}^{-1} d_\gamma\right]$$

and

$$E(Z^2) = [\det(P^p)\det\{2C_\gamma + (P^p)^{-1}\}]^{-1/2} \exp[2d_\gamma' \{2C_\gamma + (P^p)^{-1}\}^{-1} d_\gamma].$$

Taking these results together, the first claim follows.

For the second claim, we note that as $\gamma \uparrow 1$

$$C_\gamma \sim (1-\gamma) H'(I - K_1' H') R^{-1} (I - HK_1) H,$$
$$d_\gamma \sim (1-\gamma) H'(I - K_1' H') R^{-1} (I - HK_1)(y - H\mu^p)$$

because K_γ and Q_γ are continuous. The result then follows by a straightforward computation.

REFERENCES

ALSPACH, D. L. & SORENSON, H. W. (1972). Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Trans. Auto. Contr.* **17**, 439–48.

ANDERSON, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–903.

Anderson, J. L. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus* A **59**, 210–24.

Bengtsson, T., Snyder, C. & Nychka, D. (2003). Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.* **108**, 8775.

Bocquet, M., Pires, C. A. & Wu, L. (2010). Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Weather Rev.* **138**, 2997–3023.

Burgers, G., van Leeuwen, P. J. & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**, 1719–24.

Doucet, A., Godsill, S. & Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comp.* **10**, 197–208.

Dovera, L. & Della Rossa, E. (2011). Multimodal ensemble Kalman filtering using Gaussian mixture models. *Comp. Geosci.* **15**, 307–23.

Drazin, P. G. & Johnson, R. S. (1989). *Solitons: An Introduction*. Cambridge: Cambridge University Press.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–62.

Evensen, G. (2007). *Data Assimilation: The Ensemble Kalman Filter*. Berlin: Springer.

Frei, M. & Künsch, H. R. (2013). Mixture ensemble Kalman filters. *Comp. Statist. Data Anal.* **58**, 127–38.

Furrer, R. & Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Mult. Anal.* **98**, 227–55.

Gaspari, G. & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quart. J. R. Meteorol. Soc.* **125**, 723–57.

Gneiting, T., Balabdaoui, F. & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Statist. Soc.* B **69**, 243–68.

Gordon, N. J., Salmond, D. J. & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar Sig. Proces., IEE Proc. F* **140**, 107–13.

Hoteit, I., Luo, X. & Pham, D.-T. (2012). Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters. *Mon. Weather Rev.* **140**, 528–42.

Houtekamer, P. L. & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126**, 796–811.

Houtekamer, P. L. & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**, 123–37.

Künsch, H. R. (2005). Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Statist.* **33**, 1983–2021.

Lawson, W. G. & Hansen, J. A. (2004). Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Weather Rev.* **132**, 1966–81.

Lawson, W. G. & Hansen, J. A. (2005). Alignment error models and ensemble-based data assimilation. *Mon. Weather Rev.* **133**, 1687–709.

Le Gland, F., Monbet, V. & Tran, V. (2011). Large sample asymptotics for the ensemble Kalman filter. In *The Oxford Handbook of Nonlinear Filtering*, Ed. D. Crisan & B. Rozovskii, pp. 598–634. Oxford: Oxford University Press.

Lei, J. & Bickel, P. (2011). A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Mon. Weather Rev.* **139**, 3964–73.

Lei, J., Bickel, P. & Snyder, C. (2010). Comparison of ensemble Kalman filters under non-Gaussianity. *Mon. Weather Rev.* **138**, 1293–306.

Liu, J. & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, Ed. A. Doucet, N. de Freitas & N. Gordon, pp. 197–223. New York: Springer.

Liu, J. S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statist. Comp.* **6**, 113–9.

Lorenz, E. N. & Emanuel, K. A. (1998). Optimal sites for supplementary weather observations: Simulations with a small model. *J. Atmosph. Sci.* **55**, 399–414.

Mandel, J. & Beezley, J. D. (2009). An ensemble Kalman-particle predictor-corrector filter for non-Gaussian data assimilation. In *Comp. Sci. – ICCS* 2009, vol. 5545 of *Lecture Notes in Computer Science*. Berlin: Springer.

Morzfeld, M. & Chorin, A. J. (2012). Implicit particle filtering for models with partial noise, and an application to geomagnetic data assimilation. *Nonlin. Proces. Geophys.* **19**, 365–82.

Musso, C., Oudjane, N. & Le Gland, F. (2001). Improving regularised particle filters. In *Sequential Monte Carlo Methods in Practice*, Ed. A. Doucet, N. de Freitas & N. Gordon, pp. 247–71. New York: Springer.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J. & Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* A **56**, 415–28.

Papadakis, N., Mémin, E., Cuzol, A. & Gengembre, N. (2010). Data assimilation with the weighted ensemble Kalman filter. *Tellus* A **62**, 673–97.

Pitt, M. K. & Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *J. Am. Statist. Assoc.* **94**, 590–9.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10**, 462–74.

REZAIE, J. & EIDSVIK, J. (2012). Shrinked $(1 - \alpha)$ ensemble Kalman filter and α Gaussian mixture filter. *Comp. Geosci.* **16**, 837–52.

SAKOV, P. & BERTINO, L. (2011). Relation between two common localisation methods for the EnKF. *Comp. Geosci.* **15**, 225–37.

SNYDER, C., BENGTSSON, T., BICKEL, P. & ANDERSON, J. (2008). Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136**, 4629–40.

SPROTT, J. C., WILDENBERG, J. C. & AZIZI, Y. (2005). A simple spatiotemporal chaotic Lotka–Volterra model. *Chaos Solitons Fractals* **26**, 1035–43.

STORDAL, A. S., KARLSEN, H. A., NÆVDAL, G., SKAUG, H. J. & VALLÈS, B. (2011). Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Comp. Geosci.* **15**, 293–305.

SUN, A. Y., MORRIS, A. P. & MOHANTY, S. (2009). Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. *Water Resour. Res.* **45**, W07424.

TIPPETT, M. K., ANDERSON, J. L., BISHOP, C. H., HAMILL, T. M. & WHITAKER, J. S. (2003). Ensemble square root filters. *Mon. Weather Rev.* **131**, 1485–90.

VAN LEEUWEN, P. J. (2003). A variance-minimizing filter for large-scale applications. *Mon. Weather Rev.* **131**, 2071–84.

VAN LEEUWEN, P. J. (2010). Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quart. J. R. Meteorol. Soc.* **136**, 1991–9.

WHITAKER, J. S. & HAMILL, T. M. (2002). Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.* **130**, 1913–24.

ZUPANSKI, D. & ZUPANSKI, M. (2006). Model error estimation employing an ensemble data assimilation approach. *Mon. Weather Rev.* **134**, 1337–54.

[*Received July* 2012. *Revised March* 2013]